# Cloud computing

Cloud computing is a model for enabling convenient and on-demand network access to a shared pool of configurable computing resources—networks, servers, storage, applications and services—which can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud computing provides three main services: cloud applications, cloud software environments and cloud software infrastructures—often known respectively as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS).
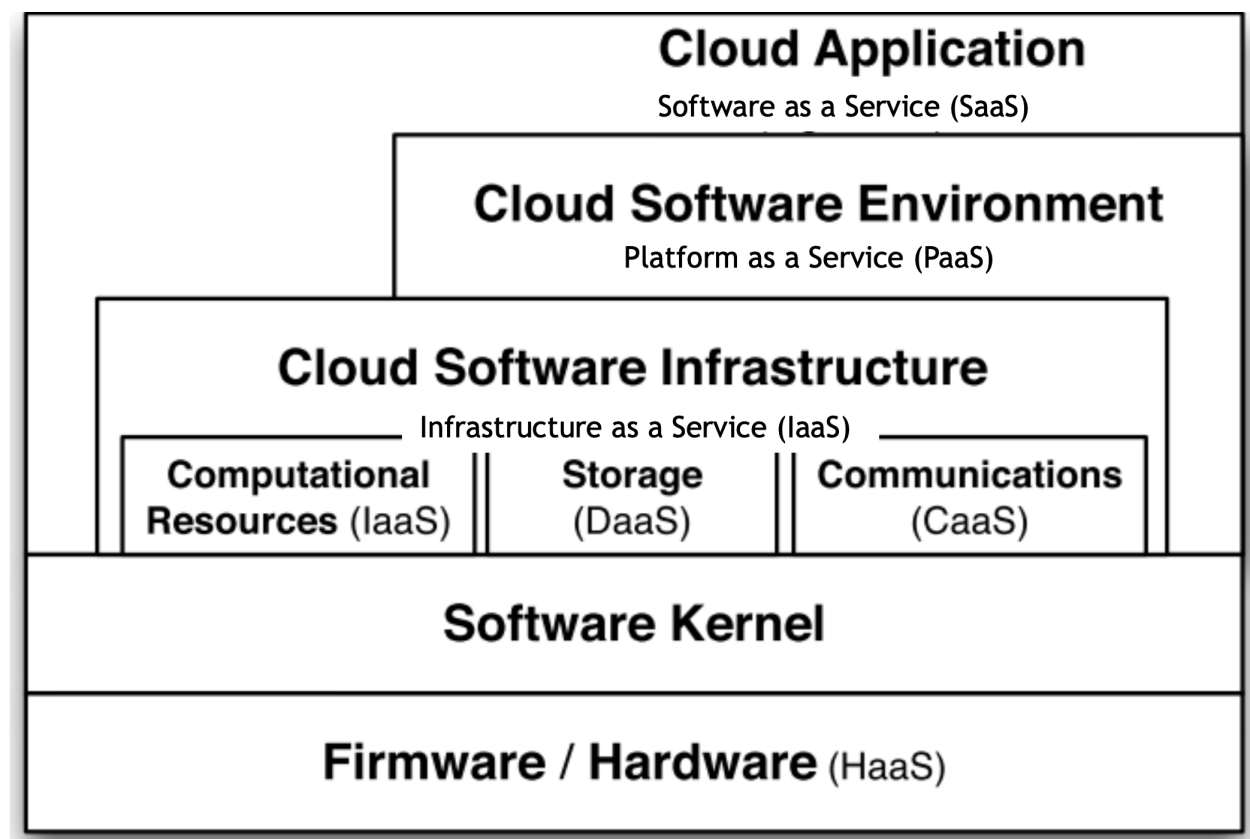
Cloud application layer: Software-as-a-Service

Users access the services provided by this layer through web portals and are sometimes required to pay fees to use them.

Cloud applications can be developed on the cloud software environments or infrastructure components.

Some examples of cloud applications are Gmail, Google Docs and SalesForce.

Cloud software environment layer: Platform-as-a-Service

The users of this layer are application developers, which providers supply with a programming language-level environment with a well-defined API in order to facilitate interaction between environment and apps, accelerate deployment and support scalability.

In the field of deep learning, some PaaS examples are Amazon SageMaker, Microsoft Azure ML and Google AI TensorFlow.

Cloud software infrastructure layer: Infrastructure-as-a-Service

The cloud infrastructure layer provides resources to the above layers. Mainly:

- Computational resources (Infrastructure-as-a-Service)
- Storage (Data-as-a-Service)
- Communications (Communications-as-a-Service)

Computational resources

Computational resources are mainly provided as virtual machines. Some providers, however, also offer dedicated hardware.

Some very popular commercial solutions for IaaS are:

- Amazon Elastic Compute Cloud (EC2)
- Microsoft Azure
- Google Compute Engine

Some open-source project also exists:

- Eucalyptus Systems
- Apache CloudStack
- Open Stack

Storage

Cloud storage solutions allow users to store their data in remote disks that they can access anytime from anywhere. This can help cloud applications reach non-functional requirements such as high dependability, fault tolerance and data consistency.

Some examples of DaaS are Dropbox, iCloud and Google Drive.

Communications

Communications are a vital component when guaranteeing Quality of Service (QoS) requirements. Different types of CaaS include VoIP and video conferencing services.

Types of clouds

There are mainly four types of clouds, depending on the use they fulfil:

- **Private** cloud: used for a single organisation; can be hosted both internally or externally to the organisation.
- **Community** cloud: a cloud used by several organisations in a shared manner; typically hosted externally—but it can be hosted by one of the organisations internally.
- **Public** cloud: cloud services that are provisioned for open use by the public; it is hosted by the organisation that provides such service.
- **Hybrid** cloud: a combination of two or more clouds of different types that remain unique entities but are bound together; this is usually hosted both internally and externally.

Private clouds

Private clouds are internally managed data centres in which the organisation sets up a virtualisation environment on its own servers.

The key benefit of a private cloud is that the organisation can control **every aspect** of the infrastructure; however, this impedes the possibility of capital investment and hinders flexibility.

Private clouds are generally useful for companies who already have a significant ongoing IT investment.

Community clouds

Community clouds are single cloud infrastructures managed by several, federated organisations. Combining together several organisations allows for economy of scale, allowing the interested parties to gain a more powerful infrastructure with lower costs.

Community clouds are technically similar to private clouds; however, a more complex accounting system is required to manage them.

Community clouds are typically hosted through the existing infrastructures of the participants to the projects; nevertheless, they can be hosted by an external organisation as well.

Public clouds

Public clouds are large-scale infrastructures available on a rental basis. They are self-service: the customers can create, manage and terminate resources via a web portal without the need for the provider to intervene.

Accountability is e-commerce based and can be either a pay-as-you-go model or a flat-rate subscription.

Hybrid clouds

Hybrid clouds are a combination of any of the previously discussed types of clouds. Often, they are used by companies that already have private cloud infrastructures but rent more compute space when sudden peaks of traffic saturate their own.

In order to simplify the development process, common interfaces, based on standards, are used by cloud providers, although no standard has already been globally accepted.

From cloud to edge computing

While cloud computing has a lot of advantages, it can also be limiting:

- It requires a constant internet connection
- It doesn't work well with low-speed connections
- Features may be limited, depending on the vendor
- It can be slow when traffic peaks
- Stored data may not be secure
- Stored data may be lost

This is where edge computing (also known as fog computing) comes into play: by moving computation nearer to the sources of data, computation can be made faster

and more reliable, easing the load on cloud infrastructures and distributing the bottleneck to more and more devices.
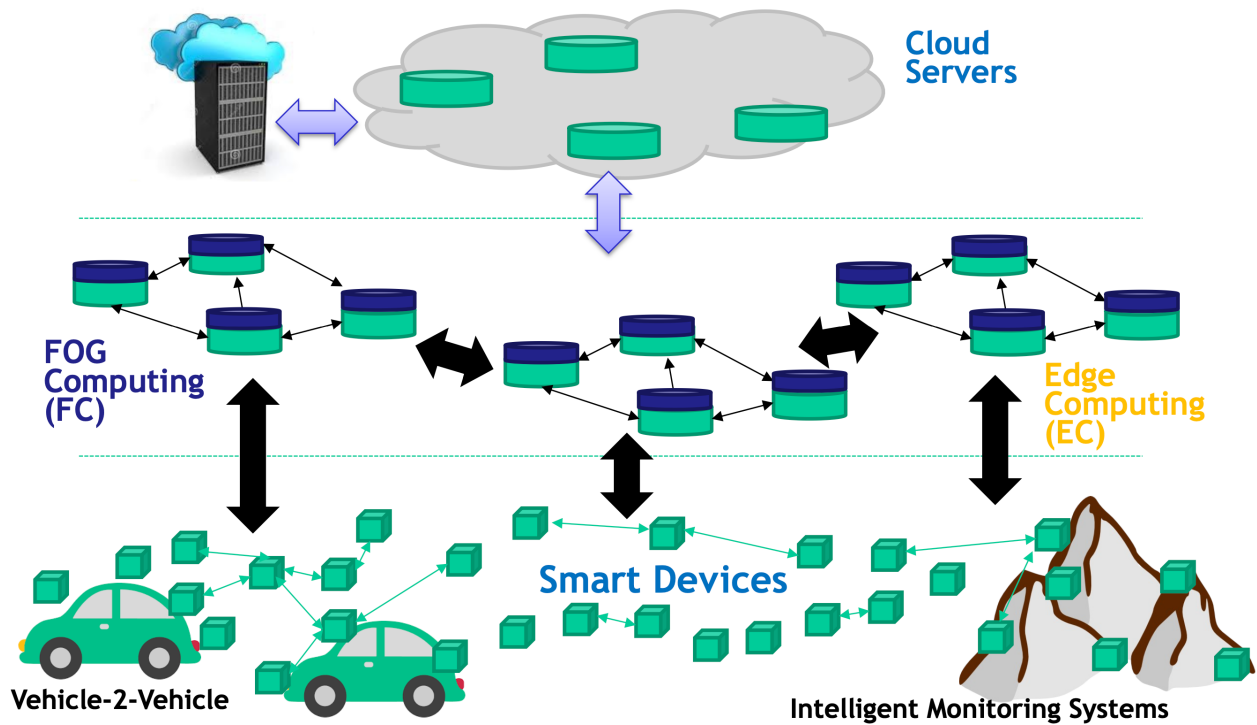
Notable applications of edge computing are Internet-of-Things (IoT), connected and autonomous cars, AI and and satellite systems.